Protein–DNA Recognition

DOI: 10.1002/anie.201101417

Protein–DNA Recognition Triggered by a DNA Conformational Switch**

*Benjamin Bouvier, Krystyna Zakrzewska, and Richard Lavery**

How DNA-binding proteins find their target sites remains a fascinating question. Early work on the Lac repressor showed that proteins specifically bind faster than simple diffusion allows. This led to the idea that recognition could be accelerated by combining diffusion with sliding along DNA and hopping between neighboring strands.[1,2] This, in turn, implied that proteins could interact with DNA in a distinct nonspecific manner. Experimental work has confirmed that proteins can indeed slide along DNA, although typical sliding distances vary from protein to protein.[3,4] Recent work also indicates that sliding follows the helical grooves of DNA.[5,6]

Crystallography of proteins bound to noncognate sites, NMR spectroscopy, and molecular simulations have all provided data on nonspecific binding, notably suggesting that proteins maintain similar orientations with respect to DNA in nonspecifically and specifically bound states (see, e.g., Refs. [7–9]). However, little is known about the transition between these states, although theoretical studies have suggested that a switching mechanism may exist, possibly involving a protein conformational change.[10]

To analyze this problem at the atomic level, we carried out a molecular dynamics (MD) study on the dissociation of a specific protein–DNA complex in water, starting from the bound conformation. We studied the sex-determining region Y (SRY) protein, which affects the gender selection in mammals and is linked to a number of gender-related pathologies.[11] The SRY protein binds in the minor groove of DNA, optimally at an (A/T)AACAAT sequence,[12] and opens the minor groove by partial intercalation of an isoleucine residue (Ile13) between two adjacent AT base pairs and induces local unwinding and bending of DNA away from the protein.[13] Using a specially designed restraint for the minimal atomic distance between any pair of non-hydrogen atoms across the protein–DNA interface (d_{MIN}), we were able to control the dissociation of the SRY protein from a 14-base-pair (bp) DNA oligomer (5′-CCTGCA-CAAACACC-3′) without biasing the conformational pathway. Using this approach, roughly 0.6 μs of umbrella sampling led to a free-energy profile for the dissociation/association process.[14] This profile showed a free-energy gain of 11.5 kcal mol^{-1} because of the SRY-DNA binding; this binding process includes passage of an energy barrier of roughly 4 kcal mol^{-1} at a separation of 4.2–3.5 Å and a secondary barrier of 2 kcal mol^{-1} at 3.1 Å (see Figure S1 in the Supporting Information). We investigated the conformational aspects of this pathway to understand the recognition mechanism.

An initial analysis showed that the conformation of the SRY protein remained remarkably stable during its separation from the DNA. Although the N- and C-terminal tails were very flexible, the three-α-helix protein core varied by a root-mean-square deviation (RMSD) of 2.2 Å at most (see Figure S1 in the Supporting Information). In contrast, the DNA oligomer underwent considerable change, linked to the SRY-induced deformations. However, our initial analysis of the separation profile also showed that many DNA conformations and protein locations occurred for a single minimal pair distance along the separation pathway. We reduced this problem by using the distance d_{AXC} from the center of the DNA helical axis to the center of mass of the core region of the SRY protein. This distance varies almost monotonically with the minimal pair distance for $d_{MIN} < 6$ Å, but increases more rapidly in the region of the main free-energy barrier (see Figure S2 in the Supporting Information). To further characterize DNA conformations we introduced two reference states: the average DNA conformation bound to the SRY protein and the average unbound B-DNA conformation, following the dissociation of the SRY protein from the DNA oligomer (see Table S1 in the Supporting Information).

Using the RMSD values with respect to the bound ($RMSD_B$) and unbound ($RMSD_U$) DNA reference states and the distance d_{AXC}, we get a clear view of what happens to DNA along the free-energy profile. Figure 1 plots conformations drawn from umbrella sampling as a function of these three variables, which are clustered (and color-coded) using a neural gas approach (see the Supporting Information). The bound state forms a tight cluster of blue points at $d_{AXC} < 13$ Å, where the bound protein stabilizes the deformed DNA conformation. This state is separated by a RMSD value of roughly 4 Å from the unbound DNA reference state, characterized by the loose cluster of yellow points at $d_{AXC} >$ 20 Å. The most interesting feature lies between these two regions (13 Å $> d_{AXC} >$ 20 Å), where the DNA conformations clearly split into two paths (green and red). The center of the two-path region occurs around $d_{AXC} = 16$ Å. Here, both paths are found at a RMSD value of around 3 Å relative to the bound DNA reference state, whereas the upper path (path 1) is located at a RMSD value of 5.2 Å relative to the unbound

[*] Dr. B. Bouvier, Dr. K. Zakrzewska, Dr. R. Lavery
Université Lyon 1/CNRS, UMR 5086, Bases Moléculaires et Structurales des Systèmes Infectieux, IBCP
7 passage du Vercors, 69367 Lyon (France)
E-mail: richard.lavery@ibcp.fr
Homepage: http://www.ibcp.fr

[**] This work was supported by the ANR Blanc ALADDIN project and by the CNRS. Computational resources were granted by the Centre Informatique National de l'Enseignement Supérieur (CINES).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201101417.

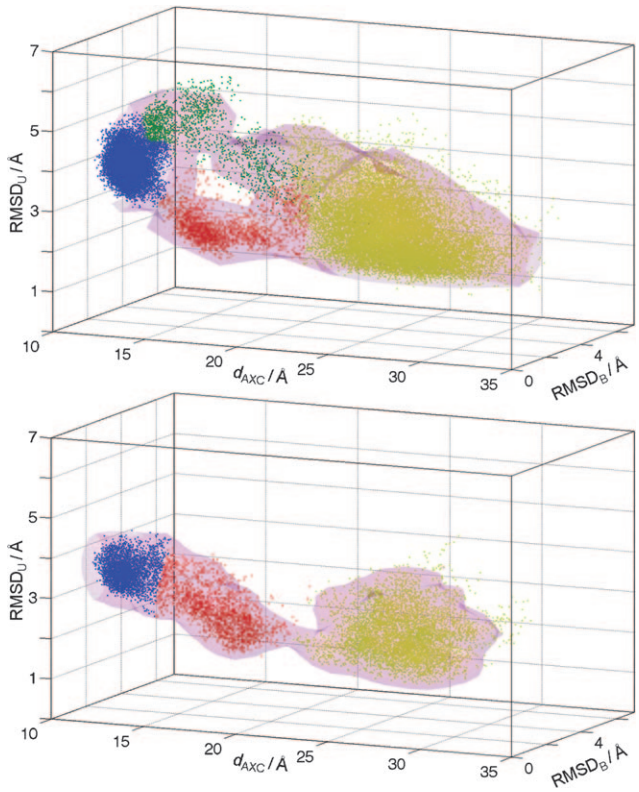


Figure 1. Representation of the conformational space for the wild-type (top) and mutated (bottom) SRY–DNA complex. Each dot represents a conformation, which is color-coded by its state (bound: blue, unbound: yellow, path 1: green, path 2: red). The purple envelope is an isodensity surface.

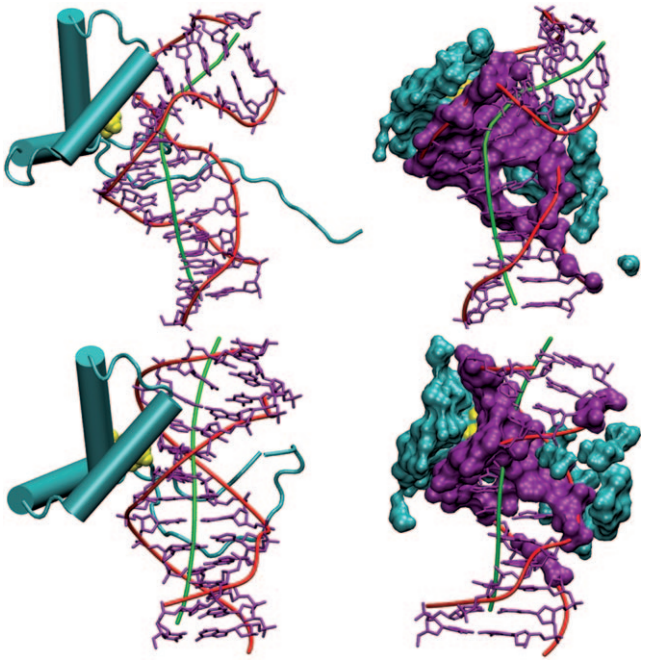


Figure 2. Representative conformations of the system (DNA: purple, protein: cyan, and Ile13: yellow) along path 1 (top) and path 2 (bottom). Left: representation that highlights the DNA axis (green tube) and backbone (red tubes). Right: surface representation of the SRY–DNA interface.

DNA reference state and the lower path (path 2) is much less perturbed with an RMSD value of only 2.2 Å.

Figure 2 shows conformations from the center of the two paths. The SRY protein has a similar orientation with respect to the DNA oligomer in both cases. When the SRY protein approaches the DNA oligomer, the protein rotates, presumably guided by electrostatic interactions, until its second α-helix is aligned with the minor groove of the DNA oligomer and its orientation stabilizes (see Figure S4 in the Supporting Information). The DNA oligomer however reacts differently to the protein along the two pathways (Table S1 in the Supporting Information). Path 1 shows a sharp kink at the A8pA9 step, where Ile 13 of SRY will finally intercalate, and also at the following AA step (with base-pair rolls of 20° and 53°, respectively), leading to an overall bend of 59°. The DNA oligomeralso has a strongly reduced twist at A8pA9 (12°) and a minor groove much wider than the major groove (11.9 Å versus 7.8 Å). In contrast, path 2 shows smaller kinks at A8pA9 and A9pA10 (base-pair rolls of 5° and 7°, respectively) and an overall bend of only 19°. The A8pA9 twist is 32° and the minor groove remains narrower than the major groove (10.7 Å versus 12.3 Å). Thus, many features of path 1 are similar to the bound state of DNA, whereas path 2 is only weakly perturbed from the unbound state.

These conformational differences cause the SRY–DNA interface to evolve differently when the protein continues to approach. Although the interface looks better packed in path 1 (Figure 2), the total protein–DNA contact area is similar in both paths (ca. 100 Å^2, see Figure S3 in the Supporting Information). However, at $d_{AXC} = 14.5$ Å water molecules are abruptly excluded from the path 1 interface and the contact area increases to 500 Å^2. A similar change occurs along path 2, but only when the SRY protein is much closer ($d_{AXC} = 13$ Å). The distance $d_{AXC} = 14.5$ Å corresponds to the top of the main free-energy barrier ($d_{MIN} = 3.75$ Å).[14] The conformations adopted along path 1 thus allow dehydration to occur earlier as the protein approaches.

We can now ask what is the most likely route to site-specific binding. Counting the transitions occurring between the regions along the free-energy pathway (Figure 3) shows that although paths 1 and 2 are similarly populated (implying a free-energy difference of < 0.1 kcal mol^{-1}), path 1 has a much higher chance of reaching the bound conformation. Path 2 is most likely to return to the unbound state, although it can also transit to path 1.

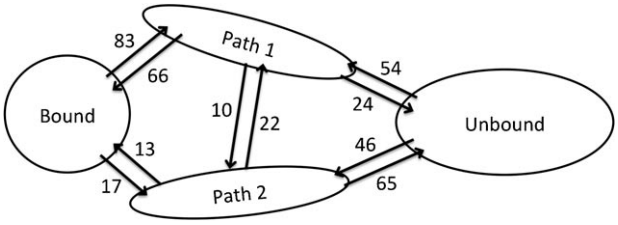


Figure 3. Percentage of transitions observed from any region of the phase space during MD simulations.

The emerging picture is that electrostatic interactions attract the protein and orient it with respect to the DNA oligomer (we cannot easily use umbrella sampling for long-range interactions, but a Poisson–Boltzmann calculation shows a favorable protein–DNA interaction of ca. 5 kcal mol^{-1} at $d_{AXC} = 30$ Å). When the protein reaches the energy barrier linked to the dehydration of the interface, it triggers the DNA oligomer to make a conformational change, which prepares the passage to the bound conformation. Analyzing the sequence-dependence of the complexation energy with the ADAPT approach[15,16] at $d_{AXC} = 16$ Å along path 1 shows that this conformational change allows the protein to partially recognize its target sequence (mostly through deformation of DNA). No such recognition occurs along path 2 (Figure 4). Following deformation of DNA, the protein can continue along path 1 to the bound state.

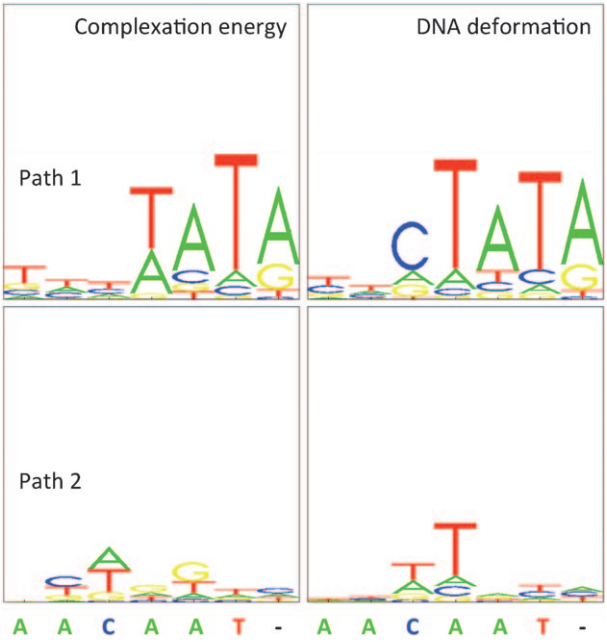


Figure 4. DNA-sequence specificity of the SRY protein shown as logos. Top: path1; bottom: path2. Left: using the total complexation energy, right: contribution from the deformation of DNA. Lower line: experimental consensus sequence.

What happens when the SRY protein approaches the wrong DNA sequence? We tested this possibility by mutating A8pA9 to G8pG9 in the bound conformation and then recalculated and analyzed the corresponding dissociation pathway. Figure S4 in the Supporting Information shows a similar free-energy profile with the mutant sequence. Because we obtained free-energies from the weighted histogram analysis method (WHAM) including an undefined constant and we cannot sufficiently sample the conformational space for the dissociated complexes, we have used nonlinear Poisson–Boltzmann electrostatic energy calculations at $d_{MIN} = 3.5$ Å to position the curves with respect to one another. These calculations imply that the mutant binds the SRY protein roughly 3 kcal mol^{-1} less well than the wild-type protein (which is in line with experimental data showing that A→G substitutions at positions 8 or 9 lead to a weaker binding of the SRY protein).[17] Figures S5 and S6 in the Supporting Information show that the SRY protein occupies a similar position with respect to the DNA oligomer at short distances, but intercalation of the Ile13 residue causes more perturbation of the DNA structure in the case of this mutant. The consequence for the DNA conformation (Figure 1) is that the bound conformation is lost as soon the SRY protein moves away from the mutant sequence. Path 1 no longer exists, and thus, with an incorrect target sequence, the protein is unable to trigger the DNA conformational switch and no recognition occurs.

This study supports the idea of an energy barrier (at a distance d_{MIN} of ca. 4 Å) linked here to the dehydration of the protein–DNA interface,[18] separating the nonspecifically and specifically bound states. A sequence-specific DNA conformational switch (rather than a protein switch)[10] controls the passage through this barrier.

Received: February 25, 2011
Revised: April 11, 2011
Published online: ■■ ■■, ■■■■

Keywords: conformational transitions · DNA · molecular dynamics · protein–DNA recognition · proteins

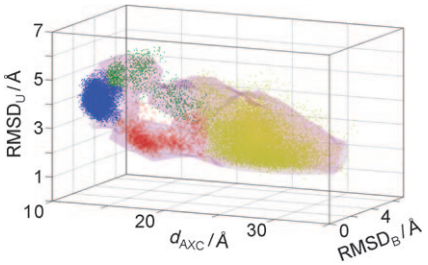
[1] R. B. Winter, O. G. Berg, P. H. von Hippel, *Biochemistry* **1981**, *20*, 6961.
[2] P. H. von Hippel, O. G. Berg, *J. Biol. Chem.* **1989**, *264*, 675.
[3] I. Bonnet, A. Biebricher, P. L. Porté, C. Loverdo, O. Bénichou, R. Voituriez, C. Escudé, W. Wende, A. Pingoud, P. Desbiolles, *Nucleic Acids Res.* **2008**, *36*, 4118.
[4] J. Gorman, E. C. Greene, *Nat. Struct. Mol. Biol.* **2008**, *15*, 768.
[5] B. Bagchi, P. C. Blainey, X. S. Xie, *J. Phys. Chem. B* **2008**, *112*, 6282.
[6] P. C. Blainey, G. Luo, S. C. Kou, W. F. Mangel, G. L. Verdine, B. Bagchi, X. S. Xie, *Nat. Struct. Mol. Biol.* **2009**, *16*, 1224.
[7] H. Viadiu, A. K. Aggarwal, *Mol. Cell* **2000**, *5*, 889.
[8] J. Iwahara, G. M. Clore, *Nature* **2006**, *440*, 1227.
[9] O. Givaty, Y. Levy, *J. Mol. Biol.* **2009**, *385*, 1087.
[10] M. Slutsky, L. A. Mirny, *Biophys. J.* **2004**, *87*, 4021.
[11] C. Y. King, M. A. Weiss, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 11990.
[12] V. R. Harley, R. Lovell-Badge, P. N. Goodfellow, *Nucleic Acids Res.* **1994**, *22*, 1500.
[13] E. C. Murphy, V. B. Zhurkin, J. M. Louis, G. Cornilescu, G. M. Clore, *J. Mol. Biol.* **2001**, *312*, 481.
[14] B. Bouvier, R. Lavery, *J. Am. Chem. Soc.* **2009**, *131*, 9864.
[15] C. Deremble, R. Lavery, K. Zakrzewska, *Comput. Phys. Commun.* **2008**, *179*, 112.
[16] K. Zakrzewska, B. Bouvier, A. Michon, C. Blanchet, R. Lavery, *Phys. Chem. Chem. Phys.* **2009**, *11*, 10712.
[17] V. R. Harley, D. I. Jackson, P. J. Hextall, J. R. Hawkins, G. D. Berkovitz, S. Sockanathan, R. Lovell-Badge, P. N. Goodfellow, *Science* **1992**, *255*, 453.
[18] V. Dahirel, F. Paillusson, M. Jardat, M. Barbi, J. M. Victor, *Phys. Rev. Lett.* **2009**, *102*, 228101.

Communications

Protein–DNA Recognition

B. Bouvier, K. Zakrzewska,
R. Lavery* _____ ■■■■–■■■■

Protein–DNA Recognition Triggered by a DNA Conformational Switch



Making the switch: The analysis of molecular dynamics simulations of the SRY-protein–DNA complex shows that, when the SRY protein approaches the correct DNA target sequence, it triggers a DNA conformational switch and allows the passage from a non-specifically bound to a specifically bound state (see picture; d_{AXC} = distance between DNA and SRY protein).